

Lliçó inaugural del curs 2021-2022

Facultat de Traducció i Ciències del Llenguatge, Universitat Pompeu Fabra

Ariadna Font Llitjós

De la lingüística computacional a una intel·ligència artificial més justa i ètica

Moltes gràcies per la invitació. Em fa molta il·lusió poder ser aquí avui amb vosaltres per explicar-vos per què val molt la pena estudiar el que tot just heu començat i aprendre llengües, així com altres camps que coneixereu durant la carrera.

Abans de començar, vull agrair molt especialment l'interpret de signes, sense ella la sessió d'avui no seria accessible a persones a les quals avui també voldria arribar.

Vicerector, degana, companyes i companys, estudiants, amics i família, per a mi és un gran honor ser aquí com a convidada en l'acte d'inauguració del curs acadèmic de Traducció i Interpretació i de Llengües Aplicades de la Universitat Pompeu Fabra.

En preparar la lliçó inaugural d'avui, vaig pensar què m'hauria agradat que m'expliquessin fa 29 anys (a l'octubre del 1992) quan jo seia on sou ara vosaltres. I després de pensar una bona estona, i d'una conversa transoceànica amb la meva estimada amiga Elisenda Bernal, aquí present, vaig veure clar que el que us voldria transmetre avui és que sovint, quan comences la carrera de Traducció o de Lingüística Aplicada, i probablement moltes altres carreres, mai saps on i quan et pots trobar el que t'apassiona i et marcarà la resta de la vida.

Crònica d'un amor no anunciat: la lingüística computacional

Avui, per començar, us explicaré la crònica d'un amor professional, que se'm va posar davant i que jo prèviament ni sabia que existia: la lingüística computacional.

Potser és oportú que abans de continuar comparteixi també les meves expectatives en entrar a la carrera de Traducció. Havia triat Traducció i Interpretació perquè tenia el somni de poder ser intèrpret a l'ONU. Sempre m'han agradat les llengües, viatjar, conèixer altres cultures i, sobretot, poder parlar amb gent d'altres llocs en la seva llengua.

Vaig tenir molta sort amb el professorat de la facultat, des de la Dra. Maria Teresa Cabré en Terminologia fins a la Dra. Isabel-Clara Simó en Literatura Catalana, i vaig gaudir amb totes les assignatures, però amb una molt especialment.

Es pot ben bé dir que em vaig enamorar de la lingüística computacional.

De la mà del Dr. Toni Badia, vaig començar aprenent Prolog, un llenguatge de programació declaratiu que permet expressar fets (com ara l'entrada d'un diccionari bilingüe) i relacions (com ara regles de traducció). Em recordava molt la lògica aristotèlica que havia estudiat a l'institut, que ja m'havia captivat en el seu moment.

Example: The goal

```
merge([subj = [spec = def,
              num = sg,
              pred = girl
              | RSubj1]
      | R1],
      [pred = hand(subj,obj2,obj),
       tense = present,
       subj = [num = sg | RSubj2]
       | R2]).
```

yields the instantiations

```
R1 = [pred = hand(subj,obj2,obj), tense = present | R2]
```

```
RSubj2 = [spec = def, pred = girl | RSubj1]
```

El fet de poder donar instruccions als ordinadors de com traduir frases bàsiques de l'anglès al català em va atraure immediatament per diversos motius. El primer, la sensació d'eufòria que t'envaeix quan, després de diversos intents fallits (normalment per alguna coma o error sintàctic) veus que finalment l'has clavat i l'ordinador, obediènt, executa les ordres a la perfecció, tal com les hi has donat.

El segon, per l'objectivitat de si ho has aconseguit o no (tan elusiu en humanitats). Al contrari de les traduccions, que mai tenen una única versió correcta, cosa que fa que la seva avaluació sigui inherentment subjectiva, quan fas córrer un programa no hi ha cap dubte sobre el veredict. És bastant binari, diguem-ne.

El tercer motiu pel qual em va captivar la lingüística computacional és que em va obrir un món de teories sobre l'origen de les paraules i de les idees, el qual és fonamental per poder avançar en l'aplicació de tecnologies de la llengua a problemes reals; en el meu cas, com fer que les

màquines puguin traduir textos no literaris (com ara les instruccions d'una eina o d'un electrodomèstic) en comptes que ho hagin de fer traductors, els quals poden preferir traduir textos més enriquidors i amb un impacte literari, social o bé científic molt més contundent.

La lingüística computacional, i el seu cosí germà, el processament del llenguatge natural,¹ són una part molt important de la intel·ligència artificial (IA).

Sigui dit de passada, jo prefereixo parlar d'intel·ligència augmentada, més que no pas d'intel·ligència artificial, en el sentit de centrar tant el disseny com el desenvolupament d'aquestes tecnologies en les persones i donar-los el control, de manera que les màquines i els programes serveixin per complementar i augmentar la intel·ligència humana.

Si em permeteu un breu parèntesi, perquè us situeu: en el moment que vaig entrar en aquest món, la intel·ligència artificial estava encara hivernant (*AI Winter*) amb alguns acadèmics i científics molt dedicats, però sense suficient capacitat computacional ni d'emmagatzematge perquè els models d'aprenentatge automàtic fossin prou potents com per poder saltar del món acadèmic al món de la indústria i a la vida quotidiana.

Des d'aproximadament el 2012, es considera que la IA va deixar d'hivernar per entrar en una fase de creixement exponencial, com ho havien fet anteriorment la capacitat d'emmagatzematge i de computació.

Tornant als meus anys d'estudiant de Traducció a la UPF, tan inesperat va ser trobar-me amb la lingüística computacional com ho va ser adonar-me que, per una d'aquelles ironies de la vida, Interpretació era l'assignatura que més em va fer patir (fins i tot més que l'alemany, i això que no en sabia abans de començar la carrera!). Per sort, en el moment clau de l'examen final d'Interpretació Simultània, el cervell em va fer un clic i vaig aprovar, però era molt clar que a l'ONU d'interpret no hi entraria.

Després d'aprendre molt i de tres estades a l'estranger (a Saint Andrews, Escòcia; a Bochum, Alemanya, i a Londres), vaig acabar fent el meu projecte de final de carrera amb el Dr. Toni Badia, titulat "*A bottom-up chart parser for PATR² unification grammars in prolog*". És a dir, vaig escriure un analitzador sintàctic en Prolog per a gramàtiques d'unificació, un formalisme en el qual es basen diverses teories lingüístiques com ara la gramàtica lexicofuncional (LFG) i la *Head-driven Phrase Structure Grammar* (HPSG).

¹ NLP takes a more pragmatic perspective and says that we wish to build systems that facilitate some language interface. The difference is that Computational Linguistics tends more towards Linguistics, and answers linguistic questions using computational tools. <https://philosophy-question.com/library/lecture/read/356039-is-computational-linguistics-nlp>

² PATR-II is a linguistic formalism used in [computational linguistics](#), developed by Stuart M. Shieber. It uses [context-free grammar](#) rules and feature constraints on these rules.

En acabar la carrera, no n'havia tingut prou i, a més d'entrar a treballar a l'Institut de Lingüística Aplicada (IULA) d'ajudant de recerca, vaig fer un doctorat interuniversitari de Ciència Cognitiva i Llenguatge que la UPF ofería conjuntament amb la UPC, la UB i la UAB, entre altres universitats.

Dos anys més tard, en acabar els cursos de doctorat, vaig rebre una beca de La Caixa per anar a fer estudis de postgrau als Estats Units. Vaig entrar a fer un màster en Tecnologia de la Llengua i de la Informació a la Universitat Carnegie Mellon, a Pittsburgh, Pennsilvània, una de les poques universitats en aquella època que tenia un departament especialitzat en lingüística computacional i processament del llenguatge natural. Carnegie Mellon University (CMU) a més era i segueix sent puntera en informàtica en general, i tot i que ja havia fet algun curs de programació, al costat dels meus companys de màster i de doctorat que programaven des de petits, jo tenia molta feina per endavant.

Van ser dos anys molt intensos, dormia poc, i quan dormia acabava somiant en codi, per intentar resoldre algun problema de programació en el qual estava encallada. La tesi de màster la vaig fer sobre la pronúncia de noms estrangers en anglès i de com en millora la pronúncia si en sabem l'origen ([Improving Pronunciation Accuracy of Proper Names with Language Origin Classes](#)). Em vaig ficar, doncs, en temes de processament de la parla amb el Dr. Alan Black, una eminència en el camp de síntesi de la parla.

En acabar el màster, vaig voler continuar amb el doctorat. (Us diré un secret: no cal cap beca per fer el doctorat als EUA. Si t'accepten, ells et paguen els cursos i un salari d'ajudant de recerca amb el qual pots viure perfectament.) Amb el director del departament, el Dr. Jaime Carbonell, com a director de tesi, vaig tornar al tema que ja m'havia captivat quan era estudiant de Traducció: la traducció automàtica, concretament, com fer servir postedicions per millorar regles de traducció de manera automàtica ([Automatic Improvement of Machine Translation Systems](#)). Si coneixeu Duolingo, la meva tesi n'és la idea base. De fet, al final de la tesi, vaig estar treballant amb el fundador de Duolingo, Luis von Ahn, precisament en aquest tema.

Durant els anys de doctorat, vaig fer recerca sobre la preservació de llengües minoritàries a través de la tecnologia, en el meu cas de sistemes de traducció automàtica. En una ocasió, el Dr. Jaime Carbonell i tres professors més em van convidar a un viatge a Temuco, Xile, perquè els fes d'interpret del castellà a l'anglès. (Per sort la interpretació consecutiva em va molt millor que la simultània.)

L'experiència em va permetre interactuar directament amb líders i educadors maputxes, i entendre els seus reptes i les seves aspiracions. Va ser una de les experiències extraordinàries que vaig tenir durant els meus anys a CMU.

Una altra d'aquestes experiències que són un regal de la vida va ser passar un estiu a Cusco, Perú, estudiant la llengua i la cultura quítxua, i, a la vegada, creant un lèxic i una gramàtica del quítxua al castellà, que van servir de base per fer un sistema de traducció automàtica.

Quan vaig tornar de Cusco, vaig defensar la meva tesi doctoral amb una panxa de set mesos de la meva primera filla, la Mar.

Després de doctorar-me, vaig entrar a una *start-up* de producte de cerca i personalització de la informació per a empreses, on vaig tenir diversos rols a més del de lingüista computacional, com ara programadora i dissenyadora de *software*. Hi vaig aprendre molt sobre desenvolupament de *software* àgil i *lean* i sobre disseny de *software* i experiència d'usuari, fins que vaig acabar liderant equips de programadors i de dissenyadors.

Cinc anys més tard, IBM ens va adquirir i vaig tenir l'oportunitat de seguir creixent com a professional i de formar-me com a líder, a la vegada que contribuïa i donava forma al programa d'IBM Design Thinking, amb el qual vam iniciar una transformació en la qual el disseny centrat en els usuaris comporta una millora significativa dels resultats de producte en l'àmbit de tota l'empresa. A més del paper d'educadora i facilitadora per a altres equips, a la vegada portava els meus propis equips de producte amb programadors i dissenyadors, que van ser dels primers a adoptar amb èxit l'IBM Design Thinking i Lean UX dins d'IBM.

El 2014, un sistema automàtic d'IBM va guanyar Jeopardy! contra els millors concursants de la història del programa (Jeopardy! és un concurs molt famós als Estats Units on els concursants reben pistes en forma de resposta i han d'endevinar les preguntes corresponents). Aquest fet històric va fer que la IA tornés a agafar impuls, sobretot pel que fa a les aplicacions al camp dels negocis. Arran d'aquesta victòria i del ressò que va tenir, IBM va crear una nova unitat de negocis, IBM Watson, centrada a portar a producció el sistema d'IA que havia guanyat a Jeopardy!, per tractar casos d'ús en àmbits com la medicina, la cadena de subministraments, i la tecnologia, entre d'altres.

Dos anys més tard, em van reclutar a IBM Research com a directora d'experiències de tecnologies emergents, i allí vaig crear un equip interdisciplinari d'investigadors, dissenyadors i programadors, encarregat de democratitzar la ciència i la tecnologia punteres. Va ser una de les experiències professionals més boniques, ja que vaig poder aportar tant la meua experiència en recerca com la meua experiència en disseny centrat en les persones i en desenvolupament de *software* àgil. Un dels projectes més emocionants va ser fer que un xip quàntic real fos accessible al núvol per a tothom des d'arreu del món. El meu equip va dissenyar, desenvolupar i posar en línia l'IBM Quantum Experience (<https://quantumexperience.ng.bluemix.net/qx/experience>), així com el millor SDK de codi obert per escriure programes i experiments per a un ordinador quàntic (<https://www.qiskit.org/>). Aquestes experiències van ser pioneres en un territori desconegut i començaven a definir un espai d'interacció completament nou amb els processadors quàntics. Va ser un projecte tan innovador que més d'un any després del llançament inicial continuava sent l'única manera d'accedir a un ordinador quàntic al núvol. I va tenir un èxit tan rotund que IBM va decidir convertir-lo en producte i començar a oferir-ne l'accés comercial.

Tornant a temes d'IA, dintre d'IBM Research vaig impulsar AI Challenges, un portal de projectes d'IA, on més de tres mil investigadors repartits en sis continents poden documentar,

compartir i prioritzar els projectes relacionats amb la IA que es duen a terme a IBM Research. A més de tenir coneixement de quins projectes estan en marxa i el seu grau de maduresa, un dels objectius del portal és impulsar col·laboracions entre equips diversos.

M'agrada molt la tecnologia, però encara m'agraden més les persones, i no és d'estranyar que l'experiència d'usuari i el disseny centrat en les persones s'acabessin convertint en una altra de les meves passions professionals. Tant a la *start-up* com a IBM, vaig constatar de primera mà el valor i l'empoderament que pot aportar el disseny centrat en les persones tant als equips com als productes de desenvolupament de *software*.

En realitat, crec que una de les maneres d'assolir un aprenentatge automàtic responsable és ampliant el disseny centrat en les persones a aquest camp. En aquest cas, hem d'entendre com afecten els algorismes i les decisions automàtiques tant els nostres usuaris com la societat en general, amb l'objectiu de fer-los justos per a tothom.

ML Platform @Twitter

Això és el que em va portar a Twitter fa poc més de dos anys, com a directora d'enginyeria de la plataforma d'aprenentatge automàtic de Twitter.

Crec fonamentalment en el poder de la connexió humana, i en la creació d'espais, tant en el món físic com en el digital, on tothom mereix ser escoltat (abans de ser jutjat).

L'aprenentatge automàtic o Machine Learning (ML) a Twitter té molts usos: des de preservar la integritat de la conversa a la plataforma, o seleccionar automàticament informes d'usuaris per a la revisió manual, fins a optimitzar algorismes i notificacions de rànquing i classificació.

Treballant per una intel·ligència artificial més justa i ètica

L'aprenentatge automàtic responsable és un camp relativament nou i que està creixent ràpidament.

Responsible ML és la versió d'IA responsable dintre de Twitter, en la qual portem treballant des de fa més d'un any, i té com a objectius: 1) responsabilitzar-nos de les decisions algorítmiques, 2) assegurar la justícia algorítmica dels models i dels sistemes d'aprenentatge automàtic, 3) ser transparent sobre aquestes decisions de manera pública, a través de Twitter, i 4) donar control a l'usuari.

L'ús tecnològic responsable inclou l'estudi d'efectes que poden tenir els algorismes al llarg del temps. ML a la plataforma de Twitter s'encarrega del rànquing de centenars de milions de piulades al dia i, a vegades, es poden colar biaixos de manera imprevisible i fer que un sistema comenci a comportar-se de manera diferent de com va ser dissenyat. Aquests biaixos

algorítmics sovint reproduïen o amplifiquen comportaments existents i acaben creant una experiència injusta, que privilegia un grup arbitrari de persones per sobre d'altres. Nosaltres volem estudiar aquests possibles canvis i fer-los servir per assegurar-nos que construïm una plataforma millor i més justa per a tothom.

Responsable ML consisteix en dues àrees de treball principals:

1. Algorismes equitatius - on ens fem responsables de les decisions algorítmiques per assegurar l'equitat i la justícia algorítmica dels resultats, sota la qual hem classificat models i sistemes per estudiar-ne el nivell de risc i d'impacte. A la vegada, fem anàlisis detallades de sistemes específics per veure si tenen biaixos, com ara la pàgina principal (*home timeline*) o la categorització automàtica de piulades en temes d'interès.
2. Donar control a l'usuari - on mirem de donar més control als usuaris, i assegurar-nos que les persones poden prendre les decisions que són importants per a elles.
 - Això ho fem a través de la recerca d'experiència d'usuari i del desenvolupament de característiques que porten a augmentar el control de l'usuari (*user agency*).
 - Aquesta àrea inclou també tasques per fer que els algorismes no siguin caixes negres proporcionant transparència sobre les decisions que hem pres, així com el perquè d'aquestes decisions (explicabilitat).

El recorregut per assolir sistemes d'aprenentatge automàtic responsables amb la comunitat al centre és col·laboratiu. Les solucions tècniques per si mateixes no poden resoldre els potencials efectes negatius de les decisions algorítmiques.

Cortex i META

Dintre de Cortex, que és l'organització central de ML a Twitter, i dintre de la plataforma de ML mateixa, tenim l'equip de META liderant aquesta tasca de fer que ML a Twitter sigui responsable. META és la sigla de ML Ethics, Transparency and Accountability, és a dir, ètica de l'aprenentatge automàtic, transparència i rendició de comptes.

L'equip de META està format per investigadors, enginyers d'aprenentatge automàtic, de privadesa i de seguretat, científics de dades, així com investigadors d'experiència d'usuaris.

Pel que fa al tipus de rols i d'experiència de l'equip, els membres de META inclouen investigadors amb formació i experiència en camps tan diversos com:

- l'aprenentatge automàtic i la justícia algorítmica de l'aprenentatge automàtic
- l'enginyeria informàtica
- les matemàtiques
- la física
- l'estadística
- les ciències polítiques
- HCI, és a dir, la interacció persona-màquina

Crec que pot ser interessant també compartir amb vosaltres una mica de la història de com es va formar META i com hem arribat on som ara.

L'equip de META es va formar de manera orgànica a principis del 2019, quan dos investigadors, un italià i una turca, van proposar crear un equip dedicat a temes d'ètica i de transparència dintre de Cortex.

Durant el primer any, META era un equip molt petit dintre del departament de recerca de Cortex, i el maig del 2020, en agafar el lideratge de l'equip, el vaig traslladar al departament de la plataforma d'aprenentatge automàtic, el qual, com us he comentat, es dedica a proporcionar les eines i la infraestructura per a l'aprenentatge automàtic dintre de Twitter.

Alguns van preguntar-se per què calia treure META del departament de recerca. I tot i que encara hem d'aprendre molt sobre temes d'aprenentatge automàtic responsable (atesa la incipiència d'aquest camp), mentre anem aprenent, jo em volia assegurar que també anem incorporant els nous coneixements a la plataforma de ML directament, de manera que puguem arribar a tot Twitter. La idea és facilitar que tots els professionals que treballen en temes de ML tinguin les eines i els coneixements per tenir en compte consideracions ètiques de manera proactiva i informada. Un exemple bàsic, però potent, és que les eines facin explícites les preguntes que s'han de formular els investigadors i els programadors mentre dissenyen i desenvolupen nous models i sistemes. Altres exemples inclouen eines d'explicabilitat com *know your data*, *model score cards*.

A mitjans de l'any passat, l'equip ja tenia el doble de membres, cosa que ens va permetre començar a explorar noves àrees de recerca i tenir més capacitat d'execució.

Va ser també aleshores quan vaig començar a compartir la nostra missió i estratègia amb el president i l'equip directiu de Twitter, així com amb el Consell Directiu. Vaig fer diverses presentacions sobre la feina que fa l'equip META, incloent-hi el full de ruta.

Això va portar que, el setembre del 2020, l'aprenentatge automàtic responsable i fiable es convertís en un dels tres objectius d'alt nivell de l'estratègia de ML de Twitter ([Objective #Trust: Earn the public's trust by taking responsibility \(via our actions\) for our algorithmic decisions](#)), i poc després, que s'establís com una de les iniciatives principals de tota l'empresa.

Ara fa un any, amb l'estratègia inicial per fer ML més responsable aprovada pel president i el seu equip directiu, vaig poder assegurar una inversió per fer créixer l'equip i vaig començar a buscar un líder amb experiència en el camp de l'ètica i la IA que no tan sols pogués portar l'equip en expansió, sinó que pogués portar també la iniciativa de RML a Twitter. El meu objectiu era trobar una persona establerta en el camp d'ètica de la IA i que fos reconeguda i respectada per la comunitat.

Era clau que el líder encaixés en l'equip, en la visió i estratègia, així com a Twitter, per així poder continuar creixent i reforçant una iniciativa tan important. Això em va portar a tenir moltes

converses amb diversos líders de la indústria i, al cap de molts mesos de buscar, vam poder convèncer la Rumman Chowdhury perquè s'unís a la causa.

El 22 de febrer d'aquest any, la Rumman es va incorporar a Twitter com a directora de META, i la resposta de la comunitat va ser aclaparadora, amb més de dues mil persones donant-li suport, i amb diverses piulades de gent dient que el fet de fitxar la Rumman havia donat de manera instantània més credibilitat a Twitter i a la nostra tasca de fer l'aprenentatge automàtic més responsable.

A l'article titulat "Twitter continua atraient el millor talent i el més divers", Anna Kramer, de Protocol, fa esment de "la ratxa dels últims dos anys en què Twitter està creant un *dream team* tècnic amb els líders més inclusius de la indústria amb experiència en el camp de l'ètica".

Un dels meus moments preferits d'aquest any va ser la celebració del Dia Internacional de la Dona el 8 de març juntament amb la Rumman i la Jutta Williams (la nostra Product Manager), inaugurant un dels nostres nous productes, anomenat *Spaces*, i oferint un espai per parlar de combatre biaixos en el camp de la IA amb la periodista Natashya Gutierrez.

Personalment, us puc dir que és molt enriquidor i emocionant formar part d'un equip de dones fortes i líders amb tant de talent.

L'equip continua creixent, i continuem mirant d'ampliar-ne la interdisciplinarietat, incorporant-hi professionals formats en temes d'ètica i filosofia, així com en altres camps complementaris.

Impacte

Passo ja al tema de l'impacte que té META dintre de Twitter i en la societat.

Actualment, estem portant a terme estudis i anàlisis detallades per detectar l'existència de danys potencials que poden ser causats pels algorismes que fa servir Twitter. Aquí podeu veure algunes de les anàlisis que ja hem dut a terme i d'altres en què estem treballant:

- Anàlisis de biaixos de gènere i de raça del nostre algorisme d'enquadrament d'imatge.
- Una anàlisi de justícia algorítmica per veure si les recomanacions que veieu a la pàgina principal de Twitter varien per subgrups racials.
- Així com també una anàlisi de recomanacions de contingut de diverses ideologies polítiques en set països, que esperem poder publicar ben aviat.

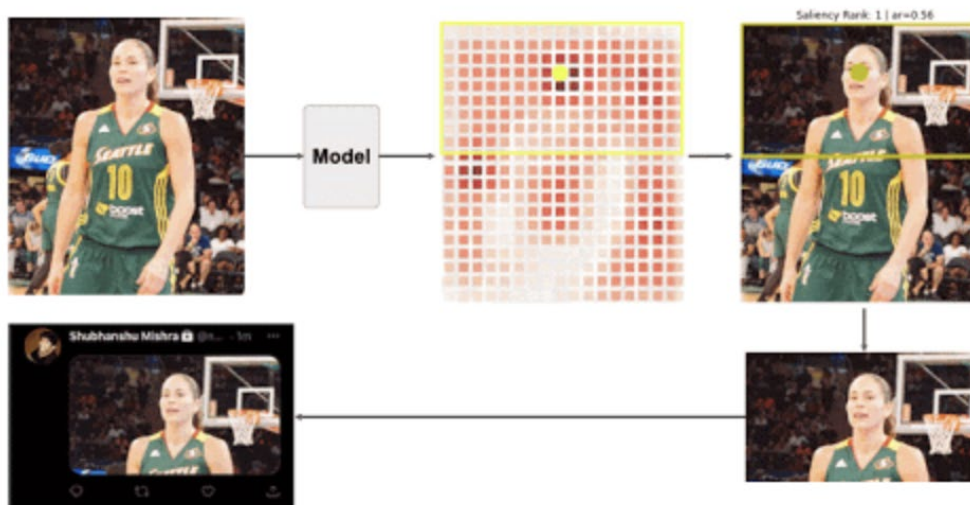
Avui voldria fer doble clic a la feina que vam fer en l'algorisme d'enquadrament d'imatge, ja que és un bon exemple del tipus de recerca que fem, i també una història d'èxit en la qual Twitter ha escoltat els usuaris i hem compartit obertament quan les decisions algorítmiques influencien negativament l'experiència d'usuari a la nostra plataforma, mostrant els resultats de la nostra anàlisi interna i responsabilitzant-nos dels algorismes, quan és necessari, fins i tot canviant el producte.

Comencem pel principi.

A finals de setembre de l'any passat, després que un usuari s'adonés que Zoom esborrava la cara d'un dels seus col·legues, a causa que l'algorisme no li reconeixia la cara pel fet de ser negre, va pujar les imatges a Twitter per compartir la seva indignació, i es va trobar que, a la previsualització que fa Twitter automàticament, també enquadrava només la seva imatge, i deixava el seu col·lega negre fora.

Això el va agafar totalment desprevingut, com podeu veure a les seves piulades sobre el tema. Molts es van sumar a l'avaluació de l'algorisme d'enquadrament de Twitter en temps real i van començar a experimentar amb altres imatges, aquesta vegada amb molt espai en blanc entre les dues imatges.

En pujar aquesta imatge, la previsualització de Twitter només mostrava la de la persona blanca, fins i tot en canviar l'ordre i la posició de les imatges.



Per posar-vos en situació, això passava en ple moviment *Black Lives Matter* als Estats Units, amb diversos casos d'assassinats de persones negres innocents per part de la policia. Total, que es va estendre com un correfoc i es va fer viral.

I com veieu amb aquest exemple, la gent no es va limitar a experimentar amb personatges reals.

Molta gent es va animar a fer tota mena de proves i la majoria no se'n sabia avenir: era racista l'algorisme de Twitter?

Twitter va introduir aquest algorisme de prominència el 2018 per enquadrar imatges ([saliency algorithm](#)), perquè totes les fotos tinguessin la mateixa mida i per permetre als usuaris veure més piulades sense haver de moure el cursor.

L'algorisme de prominència estima quina secció de la imatge volen veure primer els usuaris, per poder determinar automàticament com enquadrar la imatge i ensenyar-la de manera que sigui fàcilment visible. Els models de prominència estan entrenats en com mira l'ull humà una foto com a mètode de prioritzar el que la majoria de gent creu que és més important.

L'algorisme, entrenat amb dades de seguiment de la mirada (*human eye-tracking data*), prediu una puntuació de prominència per a totes les regions d'una imatge i tria el punt amb la puntuació més alta com a centre de l'enquadrament.

Amb la intenció de ser completament transparents, en el seu moment, també vam publicar un article tècnic i un apunt al blog descrivint tant l'algorisme com els beneficis d'aquest enfocament.

Des de la perspectiva de producte i d'experiència d'usuari, l'enquadrament d'imatges automàtic té l'avantatge que no requereix als usuaris que hagin d'editar la imatge abans de piular, cosa que es veu reflectida en les nostres mètriques d'interacció.

En aquest cas, també és important recalcar que abans de posar l'algorisme en producció, vam fer una sèrie de proves per mirar de detectar biaixos de gènere i de raça amb un conjunt de dades públiques. Els resultats inicials indicaven que no hi havia biaix racial o de gènere.

Tot i que no havíem entrenat l'algorisme amb característiques de raça ni de gènere, i que la nostra anàlisi inicial no indicava que hi hagués cap biaix, immediatament ens vam adonar que qualsevol algorisme d'enquadrament automàtic corre el risc de causar danys de representació, ja que limita la capacitat dels usuaris d'expressar-se com ells vulguin, i a més pot reforçar estereotips socials, en aquest cas la subordinació d'alguns grups d'una manera que afecta la seva identitat.

En comptes d'insistir en tecnicismes de precisió de l'algorisme, i en el fet que hi havia el 50% de probabilitat de triar una persona blanca o negra, la nostra resposta inicial va centrar-se a admetre el dany de representació que l'algorisme havia causat en un apunt al blog, on també ens vam comprometre a buscar una solució millor que donés als usuaris l'opció de triar la manera com es mostren les seves imatges a les piulades.

Cinc mesos més tard, el març passat, vam començar a [provar](#) una nova manera de visualitzar fotos a iOS i Android — sense fer servir cap algorisme de prominència.

L'objectiu, com deia, era donar més control als usuaris sobre com apareixen les seves imatges a Twitter, i a la vegada millorar també com l'altra gent veu les imatges.

Després de rebre *feedback* positiu sobre aquesta experiència, vam distribuir aquest canvi a tothom. Els canvis de producte també inclouen una previsualització de la imatge en el camp de composició de la piulada, de tal manera que els autors puguin saber com es veurà la piulada abans de publicar-la.

Aquest canvi de producte redueix la nostra dependència en l'aprenentatge automàtic per fer una funció que creiem que és millor que la dugui a terme la gent que fa servir Twitter i no pas un algorisme.

Al maig, vam publicar un altre [apunt amb les lliçons que havíem après i els resultats d'una nova anàlisi de biaixos molt més detallada](#) que havíem dut a terme.

Un resum del que vam trobar és que:

- En comparacions d'homes i dones, hi havia un 8% de diferència de paritat demogràfica a favor de les dones, i que
- En comparacions de negres i blancs, hi havia un 4% de diferència de paritat demogràfica a favor dels blancs.

Una de les conclusions principals és que no tot a la plataforma de Twitter és un bon candidat per a un algorisme, i en aquest cas, com enquadrar una imatge és una decisió que és millor que la prenguin les persones.

Com a part del nostre compromís de transparència, també vam publicar una anàlisi detallada (que es pot trobar a [ArXiv](#)) i vam obrir el nostre [codi font](#) per tal que tothom pugui reproduir i millorar la nostra anàlisi. A l'article compartim més detalls sobre els beneficis i els inconvenients de fer servir tant termes d'identitat com anotacions de tons de pell.

Considerem que aquest pot ser un full de ruta pel que fa als diversos aspectes necessaris per dur a terme un ML responsable: rendició de comptes, donant resposta a les inquietuds dels nostres usuaris i considerant l'impacte que tenen els nostres sistemes a la societat.

- Transparència, compartint la nostra auditoria i anàlisi.
- Col·laboració, obrint el nostre codi.
- Canvi de producte, quan calgui.

En resum, amb aquest incident, hem confirmat el nostre compromís de proporcionar més transparència sobre com investiguem i invertim en el fet d'entendre els danys potencials que resulten d'usar sistemes de decisió algorítmica a Twitter, tant amb les nostres accions com amb les nostres paraules.

Què ens espera en un futur pròxim?

Les aplicacions de ML responsable amb més impacte vindran de com apliquem les lliçons que anem aprenent per construir un Twitter millor i més just. L'equip de META estudia com

funcionen els nostres sistemes, i fem servir el que aprenem per millorar l'experiència de la gent a la plataforma de Twitter.

Això pot ser que requereixi un canvi de producte, com és el cas de l'enquadrament d'imatge, en què vam decidir eliminar l'algorisme i donar més control a la gent sobre les imatges que pugen a Twitter, o bé que requereixi estàndards nous que regulin les polítiques de disseny i desenvolupament quan els algorismes tenen un impacte desproporcionat per a una part de la població determinada.

Els resultats d'aquesta feina no sempre es tradueixen en canvis de producte visibles, però la intenció és que portin a generar debats rellevants sobre com desenvolupar i aplicar ML a Twitter i altres xarxes socials.

A més, continuarem compartint el que anem aprenent i les bones pràctiques per millorar el coneixement col·lectiu sobre aquest tema, tant dins com fora de Twitter, així com per ajudar-nos a millorar el nostre enfocament i retre'n compte.

Això podrà prendre la forma de recerca revisada per experts, de noves idees sobre com fer servir les dades de manera no intrusiva, descripcions sobre els nostres resultats i la nostra feina de manera entenedora per tothom, i fins i tot alguns dels nostres intents fallits de combatre aquests reptes emergents.

Continuarem treballant i col·laborant amb investigadors i entitats acadèmiques per identificar com millorar la nostra feina i rebre el seu *feedback* i consells.

El públic té un paper importantíssim en com és Twitter, i volem que aquest sigui el cas també pel que fa a un ML Responsable. El *feedback* del públic és particularment important a l'hora d'avaluar la justícia i l'equitat dels sistemes automàtics de Twitter. Quan les persones que usen Twitter formen part del procés, la presa de decisions és de més qualitat i més inclusiva; estem mirant de crear més oportunitats perquè la gent comparteixi les seves opinions sobre com hauríem d'utilitzar, o no, ML a Twitter.

Un exemple recent de com fer la comunitat participi d'aquest procés ha estat la caça de biaixos en l'algorisme de prominència ([Algorithmic Bias Bug Bounty](#)), que vam proposar a DefCon, una destinació habitual per a la comunitat de *hackers* o pirates informàtics. És la primera vegada que es feia una caça d'aquest tipus i va ser un èxit rotund, amb 31 entrades de 6 països diferents.

Per acabar, fer que els sistemes de ML siguin responsables és un viatge llarg que tot just acabem de començar. Volem explorar-lo amb un esperit obert, amb l'objectiu de contribuir al camp de l'ètica tecnològica.

Conclusió

La carrera de Traducció t'obre moltíssimes portes i moltes vegades no saps què hi ha darrere de cada porta i què és el que et pot acabar enganxant. En el meu cas, em vaig il·lusionar amb la lingüística computacional de manera totalment inesperada i imprevisible. Això m'ha donat un ventall amplíssim d'opcions en la meua vida professional, des de fer cursos de postgrau en processament del llenguatge natural als Estats Units i formar part d'una *start-up* tecnològica, fins a ser directora de tecnologies emergents a IBM Research, portant l'equip que va posar la informàtica quàntica en línia, i ara a Twitter, liderant els equips de la plataforma d'aprenentatge automàtic.

Molts potser ja ho teniu clar abans de començar, però per als qui no ho tingueu tan clar, també us pot interessar veure a què ens hem dedicat alguns dels que vam formar part de la primera promoció de Traducció i Interpretació a la UPF:

Investigació

Lingüística computacional

Lingüística aplicada

Enginyeria i programació

Disseny d'experiència d'usuari

Direcció d'equips i projectes

Direcció de departaments de llengua

Consultoria de tecnologies

Administració de sistemes

Anàlisi i suport tècnic

Traducció automàtica

Traducció i localització

Traducció científica

Traducció literària

Traducció audiovisual

Interpretació

Subtitulació

Comunicació

Redacció de continguts

Terminologia

Lexicografia

Edició, revisió, correcció i postedició

Serveis lingüístics

Documentació

Docència de llengua (catalana, castellana) o d'idiomes (anglès, francès, alemany)

Esriptura i producció d'obres de teatre

Entre moltes d'altres...

Quina és, doncs, la moral de la meva crònica d'un amor no anunciat?

Que obriu totes les portes que us posin al davant; ara és el moment d'aprendre i d'explorar. Mai se sap a quina porta podeu trobar-vos amb un amor professional que us porti molts èxits.

En el meu cas, si no hagués fet Traducció a la UPF, potser mai hauria descobert la lingüística computacional, que ha marcat la resta de la meva vida, no només en l'àmbit professional, sinó també en l'àmbit personal. I segurament ara no portaria 22 anys als Estats Units i molt menys estaria treballant a Twitter de directora de la plataforma d'aprenentatge automàtic.

Que tingueu molts èxits, i sobretot que aprengueu molt i que el vostre aprenentatge us serveixi per a la vida i per a la feina futura.

Espero també que aquest any pugueu gaudir molt de les classes presencials durant tot el curs.

Enhorabona per haver començat ja la vostra crònica professional i molt bon curs 2021-2022!

Moltes gràcies.